

University of Groningen

Functional Capacity Evaluation in Subjects with Early Osteoarthritis of Hip and/or Knee; is Two-Day Testing Needed?

van Ittersum, M. W.; Bieleman, H. J.; Reneman, M. F.; Oosterveld, F. G. J.; Groothoff, J. W.; van der Schans, C. P.

Published in:
Journal of Occupational Rehabilitation

DOI:
[10.1007/s10926-009-9179-y](https://doi.org/10.1007/s10926-009-9179-y)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2009

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

van Ittersum, M. W., Bieleman, H. J., Reneman, M. F., Oosterveld, F. G. J., Groothoff, J. W., & van der Schans, C. P. (2009). Functional Capacity Evaluation in Subjects with Early Osteoarthritis of Hip and/or Knee; is Two-Day Testing Needed? *Journal of Occupational Rehabilitation*, 19(3), 238-244.
<https://doi.org/10.1007/s10926-009-9179-y>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Functional Capacity Evaluation in Subjects with Early Osteoarthritis of Hip and/or Knee; is Two-Day Testing Needed?

M. W. van Ittersum · H. J. Bieleman ·
M. F. Reneman · F. G. J. Oosterveld ·
J. W. Groothoff · C. P. van der Schans

Published online: 15 May 2009
© Springer Science+Business Media, LLC 2009

Abstract *Introduction* The Work Well Functional Capacity Evaluation (WW FCE) is a two-day performance based test consisting of several work-related activities. Three lifting and carrying test items may be performed on both days. The objective of this study was to assess the need for repeated testing of these items in subjects with early osteoarthritis of the hip and/or the knee and to analyze sources of variation between the 2 days of measurement. *Methods* A standardized WW FCE protocol was applied, including repeated testing of lifting low, lifting overhead and carrying. Differences and associations between the 2 days were calculated using paired samples *t*-tests, intraclass correlation coefficients (ICC) and limits of agreement (LoA). Possible sources of individual variation between the 2 days were identified by Wilcoxon signed ranks tests. Pearson correlation coefficients were calculated for differences in performances between days

and differences in possible sources of variation between days. *Results* Seventy-nine subjects participated in this study, their mean (SD) age was 56.6 (4.8) years, median (min–max) WOMAC (Western Ontario and McMaster Universities) index scores for pain, stiffness and physical function were 5 (0–17), 3 (0–7) and 14 (0–49), respectively. Median (min–max) SF36 physical function was 75 (5–95), and SF36 pain score was 67 (12–76). Mean performance differences ranged from –0.2 to –0.8 kg ($P > 0.05$). ICC's ranged from 0.75 (lifting overhead) to 0.88 (lifting low). LoA were: lifting low 8.0 kg; lifting overhead 6.5 kg; carrying 9.0 kg. Pearson's correlations were low and non-significant. *Conclusions* All three tests show acceptable two-day consistency. WW FCE testing on two consecutive days is not necessary for groups of subjects with early osteoarthritis. Individual sources of variation could not be identified.

M. W. van Ittersum (✉) · C. P. van der Schans
Research and Innovation Group in Health Care and Nursing,
Hanze University Groningen, University of Applied Sciences,
P.O. Box 3109, 9701 DC Groningen, The Netherlands
e-mail: m.w.van.ittersum@pl.hanze.nl

C. P. van der Schans
e-mail: c.p.van.der.schans@pl.hanze.nl

M. W. van Ittersum · H. J. Bieleman · M. F. Reneman ·
J. W. Groothoff · C. P. van der Schans
Graduate School for Health Research SHARE, University
Medical Center Groningen, University of Groningen, Groningen,
The Netherlands

H. J. Bieleman
e-mail: h.j.bieleman@saxion.nl

M. F. Reneman
e-mail: m.reneman@cvr.umcg.nl

J. W. Groothoff
e-mail: j.w.groothoff@med.umcg.nl

H. J. Bieleman · F. G. J. Oosterveld
Expertise Center Health, Social Care and Technology, Saxion
University of Applied Sciences, Enschede, The Netherlands

F. G. J. Oosterveld
e-mail: f.g.j.oosterveld@saxion.nl

M. F. Reneman
Center for Rehabilitation, University Medical Center Groningen,
Haren, The Netherlands

J. W. Groothoff
Department of Health Sciences, University Medical Center
Groningen, University of Groningen, Groningen, The
Netherlands

Keywords Functional capacity evaluation · Osteoarthritis

Introduction

Osteoarthritis (OA) is the most common form of arthritis and a cause of long term disability among adults. It is a slowly progressive, chronic, non-inflammatory disease primarily of weight-bearing joints [1]. Risk factors for OA include age, occupations causing repetitive joint trauma, continuous overuse of joints, obesity, physical activities/participation in sports, gender and genetic factors [1, 2]. The American College of Rheumatology has developed classification criteria for OA of the knee and hip, which include clinical and radiographic aspects [3]. Clients with OA usually present with pain, morning stiffness, joint stiffness after periods of rest or inactivity, and joint crepitating [1]. OA is associated with absence from work, inability to work and poor quality of life [4–6].

The ability to perform daily activities is considered one of the most important outcome measures for patients with OA of the hip or knee [7]. To have a complete overview of patients' abilities is important for health related decisions, for example in referring to medical treatment and in return to work issues. Also for determining the outcome of clinical trials in OA a comprehensive measurement of disabilities should be used.

Use of self-reported measures is generally preferred over performance based testing, because questionnaires are mostly well-validated, less expensive and less time consuming [7–9]. However, in several studies performance based tests have demonstrated to provide complementary information on degree of disabilities. The authors of these studies recommend using both a performance based measure and a questionnaire to obtain a more comprehensive picture of the ability of the patient [10–12].

Performance based testing can be done by using functional capacity evaluations (FCEs), which are performance based batteries of tests aimed at measuring functional abilities. One of the better known FCEs is the Work Well Functional Capacity Evaluation (WW FCE). The WW FCE consists of 28 tests that measure activities such as lifting, carrying and bending [13, 14]. Psychometric properties of this FCE have been investigated in patients with chronic low back pain (CLBP) and in healthy subjects. Support was found for aspects of validity [15–17]. In patients with CLBP and in healthy subjects acceptable reliability of the WW FCE was found [18–20]. The original FCE demands testing on two consecutive days, with a total testing time of 4–5 h. Three items—lifting low, lifting overhead and carrying—may be tested twice, once on each consecutive day [14]. However, it is not clear whether this 2 day testing is necessary in patients with OA. The WW FCE will become

much more efficient when testing time can be reduced and testing on 1 day would be sufficient. To our knowledge, the need for repeated measurements of these three items has not been studied in OA before. Therefore, the objectives of this study were to investigate stability of three FCE test items (lifting low, lifting overhead, carrying) in subjects with OA on two consecutive days, to analyze consistency of individual test results, and to analyze whether pain, hip and/or knee complaints and disease severity are possible sources of individual variation between both days.

Methods

Subjects

Subjects participating in a large cohort study (cohort hip and cohort knee; CHECK [21]) were asked also to participate in this study. Inclusion criteria were: age between 45 and 65 years, pain and/or stiffness in hip and/or knee, and never or not longer than 6 months ago visited the general practitioner for these symptoms for the first time. Subjects were excluded when they had any other pathological condition that could explain the existing complaints (e.g. other rheumatic disease, previous hip or knee joint replacement, congenital dysplasia, osteochondritis dissecans, intra-articular fractures, septic arthritis, Perthes' Disease, ligament or meniscus damage, plicasyndrome, Bakers cyste) or co-morbidity that did not allow physical evaluation and/or follow-up of at least 10 years, malignancy in the last 5 years, and inability to understand the Dutch language. Participant selection methods are described extensively by Wesseling et al. (2008) [21]. Written informed consent was obtained from all participants. The local ethics committee approved the study.

Procedures

After an introduction of the FCE procedures, subjects were briefly instructed on how to perform each test. The evaluator first showed each test once. In this way, a total of 12 tests were performed on day 1 and 13 tests on day 2. The tests of the WW FCE protocol have been described elsewhere [19, 20]. The first three tests of day 1 (lifting low, lifting overhead and carrying) were repeated on the second day. The first test consisted of lifting a weight from the floor to a table at waist height for 5 times with gradually (4–5 increments) increasing amounts of weight until maximum. With lifting overhead, the ability to lift a weight from waist height to crown height was assessed, in 5 times and with increasing the amount of weight in 4 tot 5 steps. The carrying test consisted of two-handed carrying of boxed weights at waist height over 1.2 m, 5 times with 4–5 weight increments. Each test was to be performed within

Table 1 Description of the WWS FCE lifting low, lifting overhead and carrying test items performed on day 1 and day 2

FCE activity	Description	Scoring (kg)
Lifting low	5 Lifts from table to floor v.v.; 4–5 weight increments; <90 s	Maximum amount of weight
Lifting overhead	5 Lifts from table to crown height v.v.; 4–5 weight increments; <90 s	Maximum amount of weight
Carrying short two handed	5 Carries 1.2 m; waist height; 4–5 weight increments; <90 s	Maximum amount of weight

90 s (Table 1). The subjects were asked to perform to their maximum abilities. Testing of lifting or carrying items could be terminated for three reasons (whichever came first):

1. Subjects were explained that they were allowed to stop the procedures at any point if they wished to do so, for example, because of insecurity or pain;
2. A heart rate monitor was worn by the subjects throughout the test procedures. A test was terminated when the subject's heart rate met or exceeded 85% of his or her age-related maximum; and
3. The evaluator terminated testing if it became unsafe. Unsafety was defined as a situation in which the subject was not in full control of him- or herself and/or of the load.

After each test the evaluator recorded the results. Evaluator, time and place of assessment were held constant for the two consecutive FCE sessions. Each session lasted 2–3 h.

Before starting the FCE procedure subjects were asked to fill in three numerical rating scales (0–100 mm) on both days; one for pain in hip and/or knee at the moment, one for complaints of hip and/or knee at the moment, and one for disease severity at the moment.

Analyses

Data were analyzed using SPSS 13.0. Of the FCE protocol, only the three material handling tests performed on both days were analyzed in this study. Differences between tests on the 2 days on weight lifted and carried were analyzed using paired samples *t*-tests. One-way random Intraclass Correlation Coefficients (ICC) were

calculated to analyze association between day 1 and day 2. An ICC of 0.75 or more was considered as acceptable reliability [22, 23]. Stability of test results between the two consecutive days on group level was defined as: small and non statistically significant differences between the test scores on the 2 days, and ICC's of 0.75 or more. Bland and Altman analyses were performed to assess limits of agreement [24]. No criteria to interpret limits of agreement are available. Smaller limits of agreement indicate more stability because it indicates that the natural variation is small [24]. Individual performance differences between both days were expressed by calculating the % of subjects that scored better, worse or equal on day 2 compared to day 1. For the numerical rating scales Wilcoxon signed ranks tests were performed to analyze differences between the 2 days for these possible sources of variation in individual differences between the 2 days. Relationships between the day 1–day 2 differences for self-reported pain, hip and/or knee complaints and disease severity and the differences in FCE test performances were expressed with Pearson's correlation coefficients to identify if these three variables were possible sources of individual variability over both days. Variables with high and statistically significant correlation coefficients were considered indicators for sources of variation. A significance level of 0.05 was used.

Results

Seventy-nine subjects with early osteoarthritis of hip and/or knee were evaluated, of which 85% were female. Mean (SD) age of the participants was 56.6 (4.8) years, 13% of the subjects had complaints of hip, 22% complaints of knee

Table 2 The amount of weight handled maximally on both days and differences between test and retest ($n = 79$)

FCE activity	Day 1 mean kg (SD)	Day 2 mean kg (SD)	Difference mean kg (SD)	P^a	ICC	LoA	LoA % mean day 1
Lifting low	20.2 (8.9)	19.4 (8.5)	−0.8 (4.1)	0.10	0.88	8.0	40%
Lifting overhead	9.9 (4.9)	9.2 (4.2)	−0.6 (3.3)	0.10	0.75	6.5	66%
Carrying	20.4 (8.9)	20.3 (8.6)	−0.2 (4.6)	0.78	0.87	9.0	44%

Two-day reproducibility expressed in ICC

^a Paired samples *t*-tests; ICC intraclass correlation coefficient; LoA limits of agreement

and 65% of both hip and knee joints. At the start of the CHECK-study median (min–max) WOMAC (Western Ontario and McMaster Universities) index scores for pain

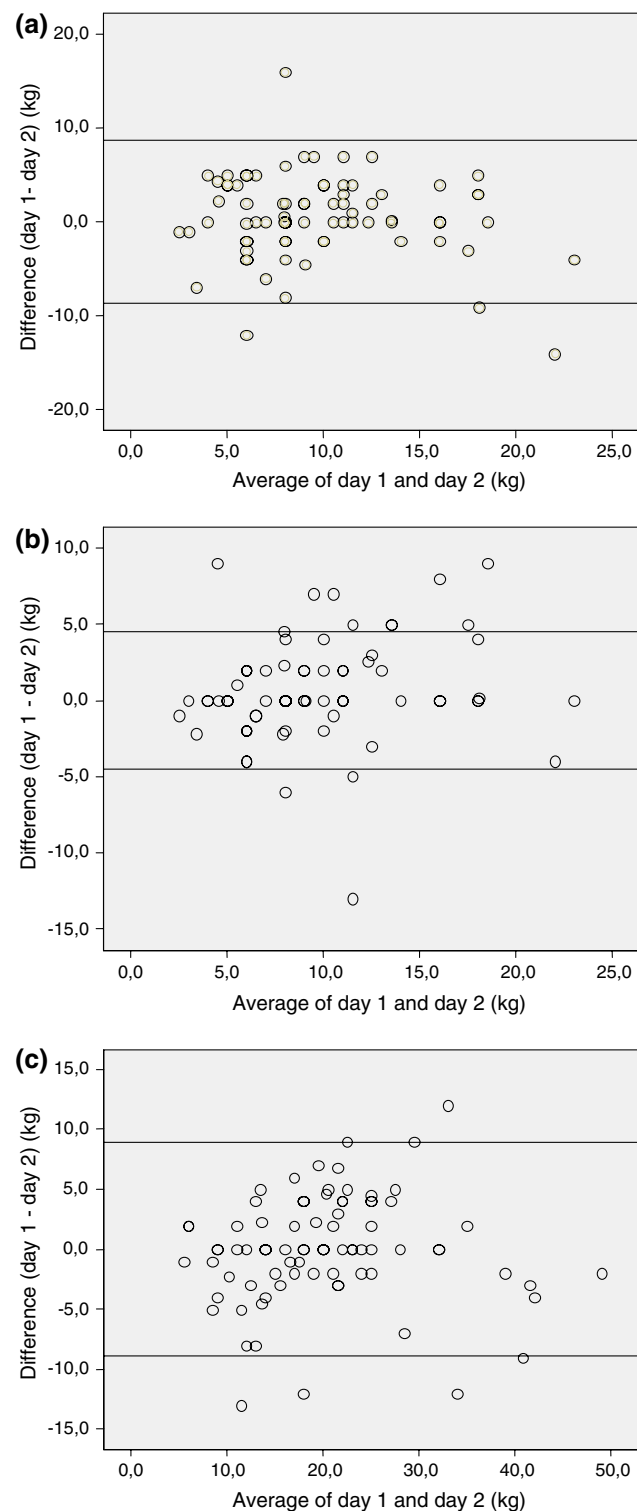


Fig. 1 Differences between the two days (day 1–day 2) plotted against average [(day 1 \pm day 2)/2] for lifting low (a), lifting overhead (b) and carrying (c) with 95% limits of agreement indicated

(range 0–20), stiffness (range 0–8) and physical function (range 0–68) were 5 (0–17), 3 (0–7) and 14 (0–49), respectively. Median (min–max) SF36 physical function was 75 (5–95), and SF36 pain score was 67 (12–76). These are comparable to the WOMAC and SF36 scores in the total CHECK cohort. In the CHECK cohort more than 65% of the participants scored Kellgren and Lawrence grade 0 for knee as well as for hip joint [21], indicating the early phase of disease in our population.

Mean (SD) scores of the 2 days for lifting low, lifting overhead and carrying on day 1 and day 2, differences between both days, ICC's and limits of agreement are presented in Table 2. Mean differences in test performance between the 2 days were statistically non-significant for all three activities ($P > 0.05$). ICC's were ≥ 0.75 for all three tests. Most tests were terminated because maximum performance was reached, only 5% of the tests were terminated when the subject was not in full control of him- or herself and/or of the load. No safety problems occurred during testing.

Bland and Altman figures are presented to analyze stability of the test results (Fig. 1). The 95% limits of agreement for lifting low are -8.8 and 7.2 , for lifting overhead 95% limits are -7.1 and 5.9 , and for carrying -9.2 and 8.8 . There were no obvious relationships between the difference between both days and their mean test scores for all three tests.

Table 3 shows the number of subjects that performed differently on the second day of testing, and reports the amount of the differences. Most individual subjects performed within a range of 20% less or more on day 2 compared to day 1, however, a large proportion of subjects performed differently on day 2. Relatively large ranges in individual performance between both days were found.

We hypothesized that the individual differences in FCE results between the two consecutive days could be influenced by pain, complaints and OA severity at the moment of the test. For this hypothesis to hold, we needed to find statistically significant differences on these variables between the 2 days, and high and statistically significant correlation coefficients between the two-day differences in these variables and the performance differences.

The self-reported pain, complaints of hip and/or knee and disease severity scores in our study population are presented in Table 4. Scores are not normally distributed, median scores on the second day are higher on all three measures, with large ranges. Differences between both days are statistically significant. On pain, 21% of subjects scored identical on both days, 14% reported less pain on the second day and 65% reported worse pain on the second day. For complaints of hip and/or knee and for self-reported disease severity similar percentages were found (21, 19, 60 and 16, 17 and 67%, respectively).

Table 3 Individual variation in FCE performance between both days

FCE activity	Equal ^a		Worse ^b		Better ^c	
	n/%	Range (kg)	n/%	Range (kg)	n/%	Range (kg)
Lifting low	49/63%	0–5	16/21%	2–16	13/17%	1–12
Lifting overhead	42/54%	0–3	21/27%	2–9	15/19%	1–13
Carrying	46/59%	0–2	17/22%	2–12	15/19%	2–13

^a Amount of weight lifted/carried on day 2 \leq 20% less or more than amount of weight lifted/carried on day 1

^b Amount of weight lifted/carried on day 2 \geq 20% less than amount of weight lifted/carried on day 1

^c Amount of weight lifted/carried on day 2 \geq 20% more than amount of weight lifted/carried on day 1

Table 4 Results for self-reported pain, complaints of hip and/or knee and disease severity (0–100) just before FCE testing on both days

Reported health problem	Day 1 [median (min–max)]	Day 2 [median (min–max)]	Difference day 1–day 2 ^a
Pain	21 (0–67)	28 (0–86)	.000*
Complaints of hip and/or knee	24 (0–73)	27 (0–90)	.000*
Disease severity	22 (0–74)	29 (0–91)	.000*

^a Based upon Wilcoxon signed ranks tests

* Statistically significant difference

Table 5 Pearson's correlation coefficients between differences in FCE performances and differences in self-reported pain, complaints and disease severity between both days

	Lifting low	Lifting overhead	Carrying
Pain	-.051	.115	-.083
Complaints of hip and/or knee	-.101	.067	-.077
Disease severity	-.004	.079	-.123

Pearson's correlation coefficients between differences in performances between days and the differences in reported health scores between both days are presented in Table 5. They were all low (<0.25) and non-significant.

Discussion

The results of this study show that 2-day consistency of lifting low, lifting overhead and carrying is sufficient, because no relevant systematic differences between test performances on day 1 and day 2 were found and all ICC's were ≥ 0.75 . As indicated by LoA, the natural variation is interpreted as large. The results of this study are similar to results of FCE studies in healthy subjects and in patients with nonspecific low back pain [18–20].

The WW FCE is one of the few to conduct testing over two consecutive days. This 2-day format is used to verify accuracy and to evaluate the effect of the first day assessment on the client [25]. Our results show that patients on

average do not perform differently on lifting and carrying on the second day of testing. Repeated testing of these three items in patients with early OA therefore may not be necessary when testing groups of subjects. Based on our results the amount of time spent on group FCE testing can be reduced.

While this may be the case for groups of subjects, in daily practice FCEs are also performed to determine capacity of individual subjects. Based on the large limits of agreement and the individual differences in FCE scores between both days found in this study, some individuals may still need retesting. Testing on 2 days might be relevant when consistency of test results over 2 days is not expected. Results of this study indicate that differences in individual test performance between two consecutive days is unrelated to changes in self-reported pain, complaints and disease severity over both days. Sources of variation for the individual performance differences between both days could not be identified in this study. Probably other variables, for example motivation or fatigue, are of importance in individual FCE test stability in subjects with early osteoarthritis. More research is needed to identify which characteristics influence individual FCE test consistency in order to be able to modify the testing procedure or to select subjects that still need 2-day testing when the FCE is used to assess physical function in individual subjects with early OA.

Former studies in FCE reliability were conducted in healthy subjects and in patients with chronic low back pain. Our sample consisted of subjects with only mild to moderate OA of hip and/or knee. Results from this study may

not apply to subjects with more severe OA and to subjects with other health conditions.

Stability of test results over 2 days covers only one aspect of the psychometric properties of a measurement instrument. Test–retest reliability of the WWS FCE in subjects with OA should also be tested with a 1–2 week time interval between test sessions. The validity of the WW FCE in OA should also be addressed in future research. Safety of the FCE in subjects with OA is another important aspect that should be further analyzed; although while in our sample the majority of the subjects seemed to experience some pain and discomfort after testing, second day performance was not significantly different from the first day, indicating that this pain increase was not related to injury or disability. During testing no safety problems occurred and no formal claims were made by the subjects.

Functional capacity evaluations test selection is based upon the job factors of the dictionary of occupational titles (DOT), a publication of the United States Department of Labor [26]. This dictionary describes the physical activities (job factors) that a job requires in a systematic way, by means of physical demands analysis. Whether the FCE is suitable for measuring one of the three, or all, main ICF health outcomes (impairment, activity limitation and participation restriction) remains unclear. The job factors described in the DOT and tested with the FCE may well be more physical demanding than activities as described in the ICF. Participation in work is an important aspect in OA because of the expected increase in prevalence of OA in working subjects and the substantial productivity related costs in OA [27, 28]. Testing of job factors could prevent productivity loss by adjustment of working place and circumstances in subjects with OA.

In conclusion, this study indicated acceptable 2-day consistency of three FCE test items in OA. The need for repeated testing of lifting low, lifting overhead and carrying on two consecutive days on group level could not be confirmed. Differences in individual test performance between both days were not related to changes in self-reported pain, complaints and disease severity over the 2 days.

References

1. Arden N, Nevitt MC. Osteoarthritis: epidemiology. *Best Pract Res Clin Rheumatol*. 2006;20(1):3–25.
2. Corti MC, Rigon C. Epidemiology of osteoarthritis: prevalence, risk factors and functional impact. *Aging Clin Exp Res*. 2003;15(5):359–63.
3. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum*. 1991;34(5):505–14.
4. Pincus T, Mitchell JM, Burkhauser RV. Substantial work disability and earnings losses in individuals less than 65 with osteoarthritis: comparisons with rheumatoid arthritis. *J Clin Epidemiol*. 1989;42:449–57.
5. Hopman-Rock M, Kraaijaat FW, Bijlsma J. Quality of life in elderly subjects with pain in hip or knee. *Qual Life Res*. 1997;6:67–76.
6. Salaffi F, Carotti M, Stancati A, Grassi W. Health-related quality of life in older adults with symptomatic hip and knee osteoarthritis: a comparison with matched healthy controls. *Aging Clin Exp Res*. 2005;17(4):255–63.
7. Bellamy N. Osteoarthritis clinical trials: candidate variables and clinimetric properties. *J Rheumatol*. 1997;24:768–78.
8. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol*. 1997;24:799–802.
9. Steultjens MP, Roorda LD, Dekker J, Bijlsma JW. Responsiveness of observational and self-report methods for assessing disability in mobility in patients with osteoarthritis. *Arthritis Rheum*. 2001;45:56–61.
10. Reneman MF, Jorritsma W, Schellekens JMH, Göeken LNH. Concurrent validity of questionnaire and performance based disability measurements in patients with chronic low back pain. *J Occup Rehabil*. 2002;12:119–29.
11. Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol*. 2006;59(2):160–7.
12. Stratford PW, Kennedy DM, Woodhouse LJ. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip and knee. *Phys Ther*. 2006;86(11):1489–96.
13. Abdel-moty E, Fishbain DA, Khalil TM, Sadek S, Cutler R, Rosomoff RS, et al. Functional capacity and residual functional capacity and their utility in measuring work capacity. *Clin J Pain*. 1993;9:2003–13.
14. King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. *Phys Ther*. 1998;78:852–66.
15. Gross DP, Battie MC. Construct validity of a kinesiophysical functional capacity evaluation administered within a worker's compensation environment. *J Occup Rehabil*. 2003;13(4):287–95.
16. Reneman MF, Jorritsma W, Dijkstra SJ, Dijkstra PU. Relationship between kinesiophobia and performance in a functional capacity evaluation. *J Occup Rehabil*. 2003;13(4):277–85.
17. Brouwer S, Dijkstra PU, Stewart RE, Goeken LN, Groothoff JW, Geertzen JH. Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. *Disabil Rehabil*. 2005;27(17):999–1005.
18. Reneman MF, Dijkstra PU, Westmaas M, Göeken LNH. Test–retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Rehabil*. 2002;12:269–75.
19. Brouwer S, Reneman MF, Dijkstra PU, Groothoff JW, Schellekens JM, Goeken LN. Test–retest reliability of the Isernhagen work systems functional capacity evaluation in patients with chronic low back pain. *J Occup Rehabil*. 2003;13(4):207–18.
20. Reneman MF, Brouwer S, Meinema A, Dijkstra PU, Geertzen JH, Groothoff JW. Test–retest reliability of the Isernhagen work systems functional capacity evaluation in healthy adults. *J Occup Rehabil*. 2004;14(4):295–305.
21. Wesseling J, Dekker J, Van den Berg WB, Bierma-Zeinstra SM, Boers M, Cats HA, et al. CHECK: Cohort Hip & Cohort Knee; similarities and differences with the OA initiative. *Ann Rheum Dis* 2008; 4. doi:10.1136/ard.2008.096164.

22. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30–46.
23. Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring quantitative variable. *Comput Biol Med*. 1989;19:61–70.
24. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;8:307–10.
25. Functional Capacity Evaluation Manual. Duluth, Minn: Isernhagen Work Systems; 1997.
26. US Department of Labor Employment and Training Administration. Dictionary of occupational titles. 4th ed. Washington, DC: US Department of Labor; 1991.
27. Leardini G, Salaffi F, Caporali R, Canesi B, Rovati L, Montanelli R, et al. Direct and indirect costs of osteoarthritis of the knee. *Clin Exp Rheumatol*. 2004;22(6):699–706.
28. Rabenda V, Manette C, Lemmens R, Mariani AM, Struvay N, Reginster JY. Direct and indirect costs attributable to osteoarthritis in active subjects. *J Rheumatol*. 2006;33(6):1152–8.